

T-test with Likert scale variables

Pedro Cosme Costa Vieira

2016

Abstract: Although Likert scale is numeric, it is intrinsically ordinal (1 – Strongly disagree to 5 - Strongly agree). Even ordinal, due to convenience it is usual to use a t-test to evaluate whether two groups are significantly different (testing population mean with unknown variance). In this paper I will investigate if when we have a survey that uses a Likert Scale, it is adequate to use a t-test. I will use bootstrapping by first “imposing” that the population verifies the null hypothesis. I conclude that, the use of the t-test it is valid to compare groups even when the variable is measured a Likert scale and the populations does not have a normal distribution.

JEL Codes: C12; C15; C83

1 - Introduction

When we ask people using a Likert scale (1- Strongly disagree to 5 – Strongly agree), for example, “Question 1 - The sensation when drinking Coca-Coke is good” and “Question 2 - The sensation when drinking Pepsi-Cola is good” and we intend to test whether there a significant difference between the two groups (the two questions, i.e., whether people prefer Coca-Cola to Pepsi Cola, H_0 : Coca-Cola = Pepsi-Cola, H_1 : Coca-Cola > Pepsi-Cola), the most simple to do it is to do a t- test on the averages (paired), *e. g.*, by using the R’s function `t.test()`.

```
t.test(Question.1, Question.2, alternative = "greater", paired = TRUE)
```

But we must remember that the Likert scale is ordinal and that the t-test is developed to use with quantitative variables that have normal distribution (Gosset, 1908).

In this paper I will investigate the pertinence of using t-test when the variable is ordinal variables (Likert Scale) and that do not have Normal distribution by using bootstrapping, random re-sampling from the sample with replacement, (Efron, 1979).

I will use a synthetic population with Q1 slightly preferable to Q2, obtained by using the next commands (R language):

```
N = 1000 #Dimension of the sample
Q.1 <- sample(c(1,2,3,4,5),N, prob=c(0.05,0.3,0.10,0.4,0.15),rep=TRUE)
Q.2 <- Q.1 + sample(c(-2,-1,0,1,2),N, prob=c(0.10,0.43,0.0,0.37,0.10),rep=TRUE)
Q.2[Question.2<1] <- 1
Q.2[Question.2>5] <- 5
```

It results a population having the following statistic characteristics (see, Table 1 and Table 2).

Scale	1	2	3	4	5	Mean-3	(>3 - <3)/N
Q1	5,0%	30,0%	10,0%	40,0%	15,0%	0,300	0,200
Q2	19,6%	10,2%	30,3%	13,2%	26,8%	0,176	0,103
Scale	-2	-1	0	+1	+2	Mean	(>0 - <0)/N
Q1-Q2	4,5%	35,4%	9,7%	43,9%	6,5%	0,124	0,104
Correlelation Q1; Q2		0,6605					

Table 1 – Distribution of the Question 1, Q1, Question 2, Q2, and the dinference Q1-Q2

		Q1					
		1	2	3	4	5	Total
Q2	1	2,7%	1,8%	0,5%	0,0%	0,0%	5,0%
	2	15,9%	0,0%	11,1%	3,0%	0,0%	30,0%
	3	1,0%	4,3%	0,0%	3,7%	1,0%	10,0%
	4	0,0%	4,0%	17,2%	0,0%	18,8%	40,0%
	5	0,0%	0,0%	1,5%	6,4%	7,0%	15,0%
Total		19,6%	10,2%	30,3%	13,2%	26,8%	

Table 2 – Joint distribution of Question 1, Q1, and Question 2, Q2

2 - The algorithm.

Step 1: Ensuring that the sample verifies the null hypothesis.

We must ensure that the re-sampling is from a “population” where the null hypothesis is observed which is done by duplicating the sample with the symmetric responses. For example, if a person responds Coca-Cola = 4 and Pepsi-Cola = 3, under the null hypothesis (the person is indifferent to the two brands), that must result from a “trembling hands” that I will correct by adding the symmetric answer, Coca-Cola = 3 and Pepsi-Cola = 4.

So if there are 1000 responses, I must add to the sample that will be re-sampled the 1000 symmetric responses.

Step 2: Re-sampling and computing the statistics.

Having the two groups (Question 1 and Question 2) I will test H_0 : Group 1 is identical to Question 2 using the i) Average and the ii) Number of people that respond $Q1 > Q2$ minus the number of people that respond $Q2 > Q1$ as statistics. I will re-sample my “corrected” sample with replacement creating a 10000 “new” sample with the same size as the original sample.

Step 3: Testing.

I will use the t-test implemented in the R- Language with the command `t.test()`, and compare it with “true values” obtained by re-sampling (10000 samples).

Step 4: Repeating.

Although $Q1 > Q2$, when I have a sample, by change it may result no-significant differences. Then, I will repeat the sample 1000 times, re-sampling each sample 10000 times (I will repeat steps 1 to 3).

#R Program used to compare the applicability of the t-test using bootstrapping

```

N <- 1000 #Number of samples
iterations <- 10000 #Number of re-samples for each sample
p.am<-0; p.am.dif<-0; p.mean<-0; p.dif<-0
for (j in 1:1000) # Repeating 1000 samples from the same population
{
#Creating the synthetic data
Q.1 <- sample(c(1,2,3,4,5),N, prob=c(0.05,0.3,0.10,0.4,0.15),rep=TRUE)
Q.2 <- Q.1 + sample(c(-2,-1,0,1,2),N, prob=c(0.10,0.43,0.0,0.37,0.10),rep=TRUE)
Q.2[Q.2<1]<-1
Q.2[Q.2>5]<-5

```

```

#p value using the values (t.test )
p.am[j]<-t.test(Q.1,Q.2,alternative = "greater",paired = TRUE)$p.value
#p value using "better" - "worst" (t.test)
d <- Q.1-Q.2
d[d<0]<- -1
d[d>0]<- 1
p.am.dif[j]<-t.test(d,(1:N)*0, alternative = "greater",paired = TRUE)$p.value
#p value using bootstrapping
# 1 – Ensuring Ho
data<-data.frame(Q1=c(Q.1,Q.2),Q2=c(Q.2,Q.1))
d<-data[1:N,1] - data[1:N,2]
# 2 - Computing the "empirical" distribution
result.mean<-0; result.dif<-0
value.to.test.mean<-mean(d)
value.to.test.dif<-(length(d[d>0])-length(d[d<0]))/N
for (i in 1:iterations)
{
  x<-sample(1:(2*N),N,rep=TRUE)
  case<-data[x,]
  result.mean[i] <-mean(case[,1]-case[,2])
  d<-case[,1]-case[,2]
  result.dif[i] <-(length(d[d>0])-length(d[d<0]))/N
}
#"empirical" p value
p.mean[j]<- length(result.mean[result.mean > value.to.test.mean])/iterations
p.dif[j] <- length(result.dif[result.dif > value.to.test.dif])/iterations
} #Close j

#Outputting results to a csv file
data.out<-data.frame(pam=p.am,pamdif=p.am.dif, pmean=p.mean,pdif=p.dif)
data.out<-data.out[data.out $pam<0.1,] #just cases that "are" significant
write.table(data.out,file = "e:/data.out.csv", dec="," ,sep=";")

```

3 - Results

My intention was to investigate two things. First, whether it is acceptable to use the t-test when the population is not normal. To evaluate this I compare the p-value that results from the t-test with the "empirical" data using as statistics the mean of the responses (See, Fig. 1) and, then, using the difference between Responses >3 – Responses <3 (See, Fig. 2).

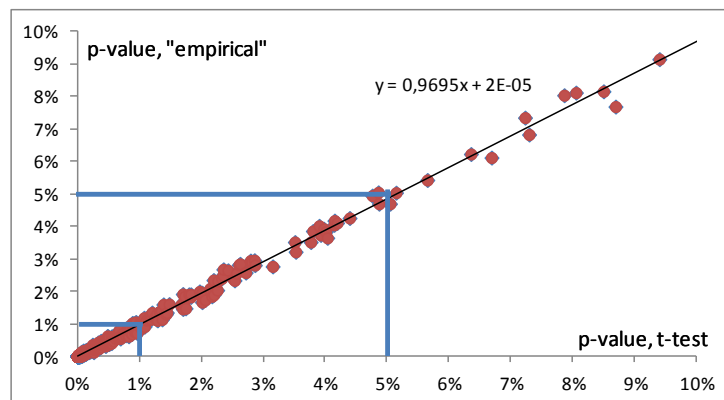


Fig. 1 – Mean - Comparing the use of the t-test with the "empirical" results (1000 points)

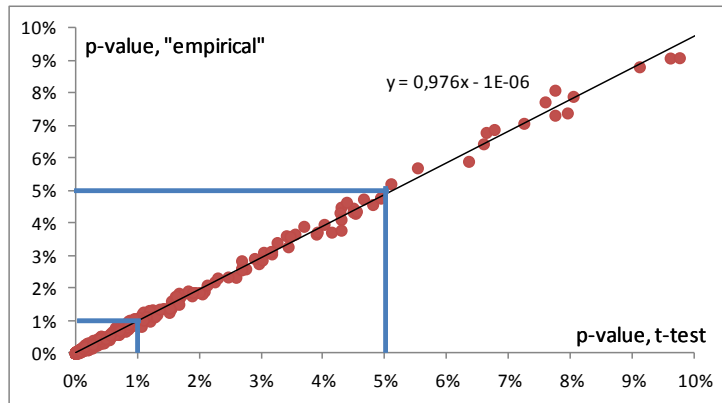


Fig. 2 – Difference - Comparing the use of the t-test with the “empirical” results (1000 points)

Due to the fact that the p-values using t-test are similar to the “empirical” p-values, I conclude that it is not a critical error to assume the results of the t-test even when the population does not have normal distribution.

The second question I intend to investigate is whether it is acceptable to use the mean of responses (Likert scale) with just the difference between those that say “Better” (>3) and those that say “Worst” (<3).

To evaluate this question I compare the p-values using the mean and using the difference using the t-teste (See, Fig. 3) and, using the “empirical” results (See, Fig. 4).

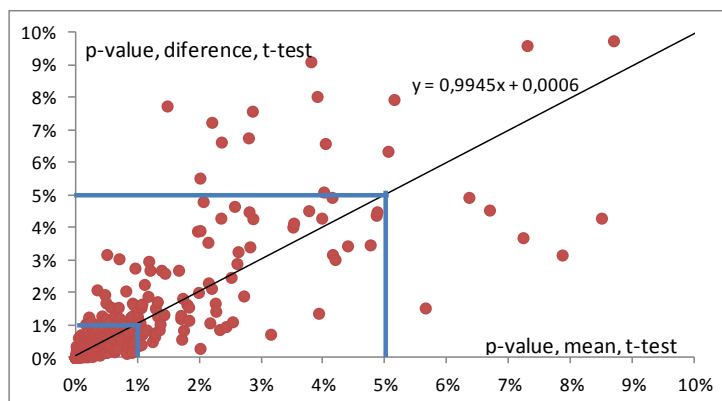


Fig. 3 –Comparing the mean with the difference (t-test, 1000 samples)

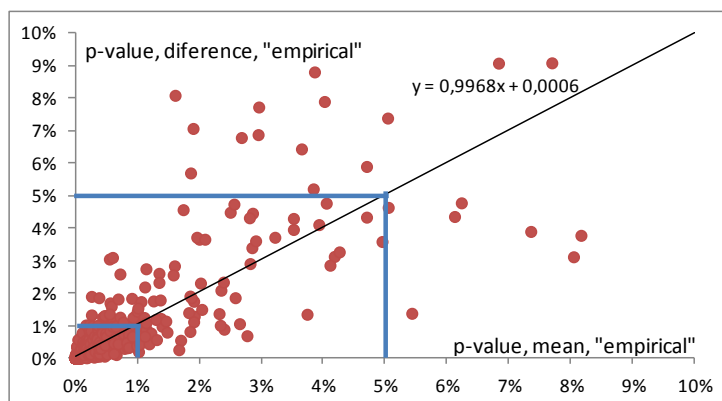


Fig. 4 –Comparing the mean with the difference (“empirical”, 1000 samples)

Although the results are not identical, using the mean we conclude from 898 out of 1000 samples that the difference is significant at 1% level and using the " >3 "-" <3 " we conclude from 892 out of 1000 samples that the difference is significant at 1% level whether in 872 samples out of 1000 the difference significant using any of the statistics. At a 5% level, we have 982 out of 1000; 978 out of 1000 and 971 out of 1000, respectively. Then, I conclude that it is statistically acceptable to test the difference of means using a t-test when the variable is a Likert scale and the populations does not have a normal distribution.

References.

Efron, B (1979). "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, 7, 1–26.

Student (1908), "The Probable Error of a Mean," *Biometrika*, 6, 1–25

<http://www.york.ac.uk/depts/maths/histstat/student.pdf>